

Object-Detection With a Varying Number of Eigenspace Projections

Michael Reiter

Pattern Recognition and Image Processing Group

Institute for Automation

Vienna University of Technology

A-1040 Vienna, Austria

rei@prip.tuwien.ac.at

Jiri Matas

CVSSP

School of EE, IT and Mathematics

University of Surrey

Guildford, Surrey GU2 5XH, United Kingdom.

G.Matas@ee.surrey.ac.uk

Abstract

We present a method allowing a significant speed-up of the eigen-detection method (detection based on principle component analysis). We derive a formula for an upper bound on the class-conditional probability (or, equivalently, a lower bound on the Mahalanobis distance) on which detection is based. Often, the lower bound of Mahalanobis distance (MD) reaches a preset threshold after computation of only a few eigen-projections. In this case the computation of MD can be immediately terminated. Regardless of the precise value of MD, the detection hypothesis (object from class Ω is detected) can be rejected. While provably obtaining results identical to the standard technique, we achieved a two to threefold speed-up in face detection experiments on images from the CMU database.

1. Introduction

Subspace methods and eigenspace decomposition (principle component decomposition) are well-suited for the target class detection problem - e.g. the task of locating a generic human face [7, 5], lips [2, 4], hands and many other structures [3] in images. In [1], Moghaddam and Pentland show how eigen-detection can be formulated as a maximum likelihood problem. Quantities commonly used in eigen-detection, the distance in feature space and distance from subspace, are interpreted as components of $P(\mathbf{x}|\Omega)$, i.e. the conditional probability density of \mathbf{x} , given that \mathbf{x} belongs to

class Ω (the class of face, hand, lip images, etc.).¹

In this paper, we make use of $P(\mathbf{x}|\Omega)$, but re-formulate the eigen-detection task in a *hypothesis testing* framework: a subimage \mathbf{x} is accepted in class Ω (i.e. a face is detected at \mathbf{x}) iff

$$P(\Omega|\mathbf{x}) \geq t_\alpha \quad (1)$$

where t_α is a threshold on $P(\Omega|\mathbf{x})$, the probability of \mathbf{x} being an instance of class Ω ($P(\Omega|\mathbf{x})$ is proportional to $P(\mathbf{x}|\Omega)$ if the probability of the non-face class is assumed constant). t_α corresponding to a chosen significance level α .

We will show that to reject the 'subimage \mathbf{x} in class Ω ' hypothesis it is often sufficient to compute only a *fraction* of the full set of eigen-projections. The proposed improved algorithm produces provably identical results to the standard method, by establishing an upper bound on $P(\Omega|\mathbf{x})$, denoted $P_i^\uparrow(\Omega|\mathbf{x})$, that is iteratively updated after every eigen-projection². If

$$t_\alpha \leq P_i^\uparrow(\Omega|\mathbf{x}),$$

which implies

$$t_\alpha \leq P(\Omega|\mathbf{x}),$$

i.e. the upper bound reaches the pre-selected threshold t_α , the hypothesis that subimage \mathbf{x} belongs to class Ω can be rejected, without computing the remaining eigen-projections.

¹for simplicity, and since all our experiments are performed on facial images, we will from now on identify the class Ω with the class of images of faces.

²the subscript i makes explicit the dependence of $P_i^\uparrow(\Omega|\mathbf{x})$ on the number of eigen-projections computed

The reduction in eigen-projection computations gained by the proposed approach depends on the number of eigenvectors in the class model and the data in which instances are searched; For certain thresholds t_α and some subimages only a single eigen-projection was sufficient to reject the ‘face (hand, lips) present’ hypothesis. In this case, eigen-detection and correlation with a single template (the mean of the training set) are equivalent; the method described below may be therefore used to detect this equivalence and is suitable for assessing the necessary complexity of a class model for a given background.

The proposed algorithm is not a new target detection method, rather it shows how to efficiently implement the eigen-detection approach, with a performance guaranteed to be identical to the standard technique of computing a fixed number of eigen-projections at every chosen location. The benefit is independent of the practically important issue of search strategy, i.e. the problem of choosing suitable subimages \mathbf{x} , i.e. locations, scales, rotation angles etc. at which the likelihood $P(\Omega|\mathbf{x})$ is evaluated.

The rest of the paper is organized as follows. In section 2 we define $P_i^\dagger(\Omega|\mathbf{x})$ and show it is indeed an upper bound of $P(\Omega|\mathbf{x})$. In section 3 we report test results on representative images - a group photo, a simple background head-and-shoulders scene and a complex background head-and-shoulders scene. We summarize the contribution in section 4.

2. Lower bound for the Mahalanobis distance

Given a training set $T_x = \{\mathbf{x}_1, \dots, \mathbf{x}_{N_T}\}$ of N -dimensional image vectors the we can estimate the mean $\mu(T_x)$ and covariance matrix Σ of the underlying distribution. We calculate by eigen-decomposition of Σ the matrix of eigenvectors Φ of Σ and the diagonal matrix Λ containing the corresponding eigenvalues in decreasing order. Φ is the basis of the Karhunen-Loeve Transform. We assume that the class-conditional probability density for \mathbf{x} is Gaussian (which is reasonable for our data [1]):

$$P(\mathbf{x}|\Omega) = \frac{e^{-\frac{1}{2}\bar{\mathbf{x}}'\Sigma^{-1}\bar{\mathbf{x}}}}{(2\pi)^{\frac{N}{2}}|\sigma|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}\mathbf{y}'\Lambda^{-1}\mathbf{y}}}{(2\pi)^{\frac{N}{2}}\prod_{i=1}^M\lambda_i^{\frac{1}{2}}} \quad (2)$$

The distance function $D_{\mathcal{M}} = \mathbf{y}^t\Lambda\mathbf{y} = \sum_{i=1}^N\frac{y_i^2}{\lambda_i}$ is called the *Mahalanobis distance*(MD). The MD is proportional to the negative logarithm of the class-conditional probability

$$D_{\mathcal{M}} \approx -\log P(\mathbf{x}|\Omega) \quad (3)$$

By keeping only the first M eigenvectors (eigenvector with the largest eigenvalues) one can obtain a *principal subspace* of the space of all possible images which captures the major correlations of the training set T_x .

Since the Mahalanobis distance is in a one to one correspondence with the class-conditional probability $P(\mathbf{x}|\Omega)$ we can use a lower bound on $D_{\mathcal{M}}$, denoted $D_{\mathcal{M}}^\downarrow(x)$, instead of an upper bound on $P(\Omega|\mathbf{x})$. It is easy to show that the Mahalanobis distance $D_{\mathcal{M}}$ of a vector \mathbf{x}_k at every step of the eigen-projection (k denotes the number of eigen-components that have already been used in the projection) becomes minimal if the vector is aligned with the eigenvector e_{k+1} that corresponds to the largest eigenvalue in the space of remaining eigenvectors $\{\Phi_i\}_{i=k+1}^N$. Thus a lower bound for the Mahalanobis distance given the projections y_k for $k < M$ onto the first k principal axis e_k is given by

$$D_{\mathcal{M}}^\downarrow(x) = \sum_{i=1}^k \frac{y_i^2}{\lambda_i} + \frac{\epsilon(\mathbf{x})_k^2}{\lambda_{k+1}} \leq D_{\mathcal{M}}(\mathbf{x}) \quad (4)$$

where $\epsilon(\mathbf{x})_k^2 = \epsilon(\mathbf{x})^2 - \sum_{i=1}^k y_i^2$ is the remaining Euclidean distance of the vector \mathbf{x} after k projections. The improved algorithm for computing the eigenspace-projections looks as follows:

Algorithm 1 Algorithm for eigen-detection

- 1: $\epsilon(\mathbf{x})_0^2 = \|\mathbf{x}\|^2, k = 1$
 - 2: $D_{\mathcal{M}}^\downarrow(x) = \epsilon(\mathbf{x})_0^2$
 - 3: **while** NOT $D_{\mathcal{M}}^\downarrow(x) \geq \tilde{t}_\alpha$ **do**
 - 4: **if** $k > \max$ **then** TARGET DETECTED
 - 5: $y_k = e_k \mathbf{x}, \epsilon(\mathbf{x})_k = \epsilon(\mathbf{x})_{k-1} - \|y_k\|$
 - 6: $D_{\mathcal{M}}^\downarrow(x) = \sum_{i=1}^k \frac{y_i^2}{\lambda_i} + \frac{\epsilon(\mathbf{x})_k^2}{\lambda_{k+1}}$
 - 7: $k = k + 1$
 - 8: **end while**
-

\tilde{t}_α is the threshold for $D_{\mathcal{M}}^\downarrow(x)$ corresponding to t_α for $P_i^\dagger(\Omega|\mathbf{x})$.

3. Experiments

Training data. A model of the face image class used in our experiments was calculated by principle component analysis of the covariance matrix of a collection of images from the publicly available M2VTS database [6]. Selected images (148 frontal views of 37 individuals, four images per person) were geometrically and photometrically registered.

Test Images. The test images were taken from the CMU Database. Since our improvement of the eigen-detection method does not depend on the search strategy for the correct scale, rotation and location we simplified the experiments by scaling the test images manually roughly to the scale of the training data. The detection was then performed at every location of the test image at a single scale.

Results. Figure 1(a) and figure shows the detection of faces on one of the test images ‘‘cnn2600’’ and the corre-



Figure 1. Detection of the face in “cnn2600” (a) and number of eigenvector projections (b)

sponding map (b) for the number of eigenvector projections that are needed for the MD lower bound to be greater than \tilde{t}_α . The grey-level codes the number of projections (black 1-2, 25% gray 3-5, 50% 6-10, 75% 11-19, white 20). At non-face regions the process is stopped early while in face regions all eigenspace projections have to be computed. A histogram for the number of projections is depicted in table 1. The number of eigen-images used for representation divided by the average number of projections defines the speedup (in this case we are using the first 20 eigen-images).

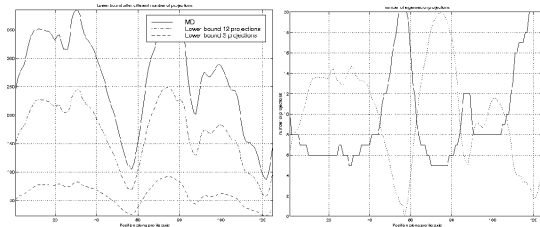
Figure 2 (b) shows the lower bound for the MD along a profile axis in the “ds9” image which goes through the center of two faces (see 2 (a)). The corresponding number of projections and the MD estimation is shown in Figure 2 (c) and (d) respectively. A scaled estimated MD-curve is also drawn over (a). At regions with low MD the number of eigenspace-projections is high.

Image	1-2	3-5	6-10	10-20	μ	speedup
ds9	0.13	0.64	0.16	0.14	7.9	2.5
cnn2600	0.12	0.62	0.17	0.17	8.0	2.5
cnn2111	0.19	0.64	0.14	0.03	7.2	2.8

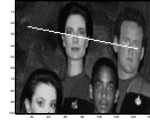
Table 1. Histogram for the number of eigenspace projections

4. Conclusions

We presented a method allowing a significant speed-up of the eigen-detection method (detection based on principle component analysis). We derived a formula for an upper bound on the class-conditional probability (or, equivalently, a lower bound on the Mahalanobis distance) on which detection is based. While provably obtaining results identical to the standard technique, we reported in section 3 a two to threefold speed-up in face detection experiments on images from the CMU database.



(a) Lower bound for MD for different size of sub-space (b) Number of Eigenspace projections



(c) Profile line

Figure 2. (a) Lower bound for the Mahalanobis distance using a different number of eigenvector projections, (b) Number of eigenspace projections, (c) Profile line though “ds9”

References

- [1] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. In S. Nayar and T. Poggio, editors, *Early Visual Learning*, pages 99–130. Oxford University Press, 1996.
- [2] C. Bregler and Y. Konig. ‘Eigenlips’ for robust speech recognition. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 2, pages 669–672. IEEE, Piscataway, NJ, USA, 1994.
- [3] T. F. Cootes and C. J. Taylor. Active shape model search using local grey-level models: A quantitative evaluation. In J. Illingworth, editor, *British Machine Vision Conference*, pages 639–648. BMVA Press, 1993.
- [4] J. Luettin, N. A. Thacker, and W. Beet. Active shape models for visual speech feature extraction. Technical report, University of Sheffield, Sheffield, UK, 1995.
- [5] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 84–91. IEEE, Los Alamitos, CA, USA, 1994.
- [6] S. Pigeon. The M2VTS database. Technical report, Laboratoire de Télécommunications et Télé-détection, Université catholique de Louvain, Louvain-La-Neuve, Belgium, <http://www.tele.ucl.ac.be/M2VTS>, 1996.
- [7] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.